Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Structure of *Mycobacterium tuberculosis* Rv2714, a representative of a duplicated gene family in Actinobacteria

Martin Graña,[a]‡ Marco Bellinzoni,[a] Isabelle Miras,[b] Cedric Fiez-Vandal,[b]§ Ahmed Haouz,[b] William Shepard,[b] Alejandro Buschiazzo[a]‡ and Pedro M. Alzari[a]*

[a]Institut Pasteur, Unité de Biochimie Structurale, URA CNRS 2185, 25 Rue du Dr Roux, 75724 Paris, France, and [b]Institut Pasteur, Plate-forme de Cristallogenèse et Diffraction des Rayons X, 25 Rue du Dr Roux, 75724 Paris, France

‡ Present address: Institut Pasteur de Montevideo, Mataojo 2020, 11400 Montevideo, Uruguay.
§ Present address: Karolinska Institutet, Department of Medical Biochemistry and Biophysics (MBB), Scheeles väg 2, 17177 Stockholm, Sweden.

Correspondence e-mail: alzari@pasteur.fr

Received 14 July 2009
Accepted 1 September 2009

PDB Reference: Rv2714, 2wam, r2wamsf.

The gene Rv2714 from *Mycobacterium tuberculosis*, which codes for a hypothetical protein of unknown function, is a representative member of a gene family that is largely confined to the order Actinomycetales of Actinobacteria. Sequence analysis indicates the presence of two paralogous genes in most mycobacterial genomes and suggests that gene duplication was an ancient event in bacterial evolution. The crystal structure of Rv2714 has been determined at 2.6 Å resolution, revealing a trimer in which the topology of the protomer core is similar to that observed in a functionally diverse set of enzymes, including purine nucleoside phosphorylases, some carboxypeptidases, bacterial peptidyl-tRNA hydrolases and even the plastidic form of an intron splicing factor. However, some structural elements, such as a β-hairpin insertion involved in protein oligomerization and a C-terminal α-helical domain that serves as a lid to the putative substrate-binding (or ligand-binding) site, are only found in Rv2714 bacterial homologues and represent specific signatures of this protein family.

1. Introduction

Mycobacterium tuberculosis infects a third of the world population and the increasing spread of multidrug-resistant and extensively drug-resistant strains is a threat to public health worldwide. The validation of new potential targets for therapeutic intervention is thus a priority in the fight against tuberculosis. Several genes annotated as hypothetical proteins of unknown function in the *M. tuberculosis* genome are now known to be confined to the Actinomycetales class or even only to the *Mycobacterium* genus and thus might represent potential candidates for specific antituberculosis drug targets (Cole, 2002). Here, we show that Rv2714 (annotated as a 'conserved hypothetical protein rich in alanine and leucine'; http://genolist.pasteur.fr/TubercuList/) is a representative of a new nucleoside phosphorylase-like family of duplicated actinobacterial proteins and we discuss the structural and functional implications of these observations.

2. Materials and methods

The Rv2714 gene encoding the whole sequence of the protein (residues 2–324) was cloned into the expression vector pDEST17 following the Gateway procedure (Invitrogen) and the resulting expression plasmid was used to transform *Escherichia coli* BL21(DE3)pLysS cells. The cells were grown in 500 ml LB medium containing 100 µg ml^{-1} ampicillin and 25 µg ml^{-1} chloramphenicol for 3.5 h at 303 K before induction with 1 m*M* isopropyl β-D-1-thio-galactopyranoside (IPTG). After 1.5 h, the cells were harvested by centrifugation, resuspended and frozen in 50 m*M* Tris–HCl buffer containing 150 m*M* NaCl and 10 m*M* imidazole pH 8.0. The cellular suspension was thawed and the cells were lysed in a French press at 97 MPa. After centrifugation, the His_6-tagged recombinant protein was purified by metal-affinity chromatography on Ni^{2+}-chelating columns (HisTrap, GE Healthcare) followed by size-exclusion chromatography (Superdex 75 16/60, GE Healthcare) and concentrated to 8.5 mg ml^{-1} for crystallization. Selenomethionine-labelled (SeMet)

protein was produced in *E. coli* B834(DE3) cells (Novagen) as described previously (Wingfield, 2000) and purified as above.

Initial crystallization screenings were carried out with sitting drops (200 nl protein solution plus 200 nl crystallization solution equilibrated against 150 µl well solution) using a Cartesian Technologies nanolitre dispenser workstation. Hits were manually reproduced and optimized. Two different crystal forms of the protein were obtained (Table 1). Hexagonal crystals (space group $P6_3$, displaying merohedral twinning) were grown in 1 *M* sodium citrate, 0.1 *M* HEPES pH 7.5, 50 m*M* ammonium sulfate and trigonal crystals (space group $P3_221$) were grown in 1.5 *M* ammonium sulfate, 0.1 *M* MES pH 6.5, 10% 1,4-dioxane.

X-ray diffraction data were collected at the ESRF synchrotron (Grenoble, France) using crystals frozen (100 K) in mother liquor plus 25%(*v*/*v*) glycerol. A 2.7 Å resolution data set was collected on beamline ID14.4 from a single SeMet hexagonal crystal at the X-ray energy corresponding to the peak of the Se *K* edge (12.661 keV, λ = 0.9793 Å), as measured by a fluorescence scan, and a 2.6 Å data set for the native protein in space group $P3_221$ was collected on beamline ID14.3 (λ = 0.931 Å). In both cases, data reduction (see Table 1) was carried out using the *XDS*/*XSCALE* package (Kabsch, 1993). For the hexagonal crystal form, a total of eight selenium sites were located with the program *SHELXD* (Sheldrick, 2008) and single-wavelength anomalous diffraction (SAD) phasing and density modification were carried out with the program *autoSHARP* (Vonrhein *et al.*, 2007). Despite the presence of merohedral twinning, the resulting electron-density map (phasing power 3.44, figure of merit 0.62) was interpretable and allowed manual tracing of about 80% of the polypeptide chain with *O* (Jones *et al.*, 1991) and *Coot* (Emsley & Cowtan, 2004). This partial model, which was initially refined with *REFMAC* (Murshudov *et al.*, 1999) from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) and then with *SHELXL* (Sheldrick, 2008) applying twin refinement (twin fraction 0.43), was subsequently used as search probe to determine the structure of the trigonal crystal form by molecular-replacement methods with the program *Phaser* (McCoy *et al.*, 2007). The new model was manually extended with *Coot*, refined with *PHENIX* (Afonine *et al.*, 2005) and validated using the *MolProbity* server (http://molprobity.biochem.duke.edu). Final refinement parameters are shown in Table 1. Atomic coordinates and structure factors of Rv2714 in space group $P3_221$ have been deposited in the Protein Data Bank with accession code 2wam (the coordinates of the partial model in space group $P6_3$ were not deposited, but are available from the authors upon request).

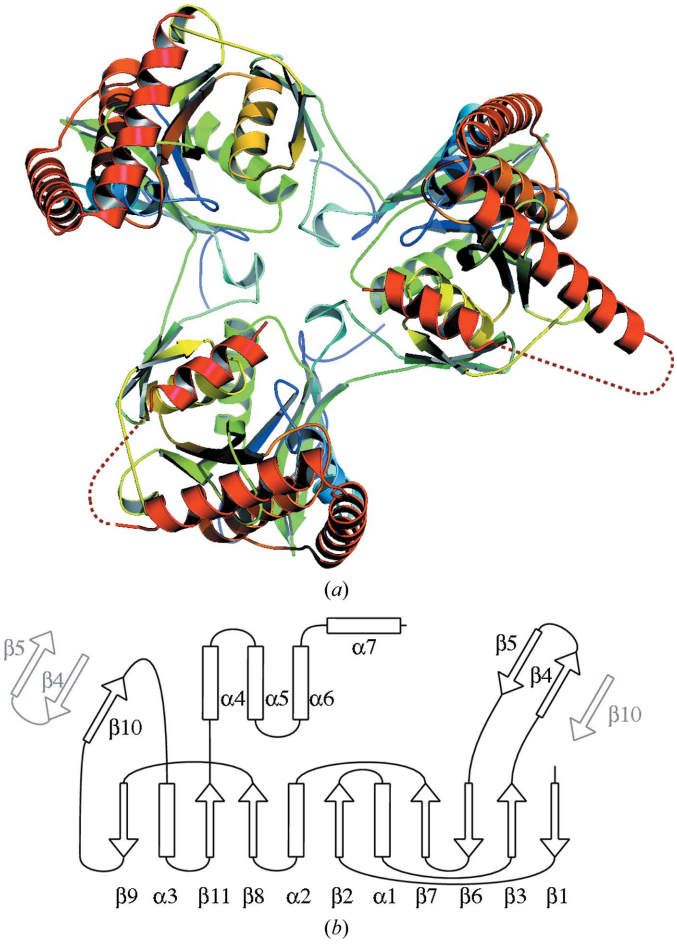

Figure 1
Structure of Rv2714. (*a*) Cartoon representation of the Rv2714 trimer, with each monomer coloured from blue (N-terminus) to red (C-terminus). The C-terminal helix α7 is only observed in two of the three monomers. (*b*) Secondary-structure topology of the Rv2714 monomer. The hairpin β4–β5 from one monomer interacts with strand β10 from a neighbouring monomer in the trimer.

3. Results and discussion

3.1. Global fold

The structure of Rv2714 has been solved in two distinct crystalline forms. The first crystallization hits led to merohedrally twinned crystals with apparent point-group symmetry 622. Nevertheless, an interpretable electron-density map could be obtained from SAD phasing using an SeMet protein crystal, which showed two monomers in the asymmetric unit and allowed the manual tracing of about 80% of both polypeptide chains. After preliminary refinement, this partial model was used as a search probe to determine the structure of the second (trigonal $P3_221$) crystal form using molecular-replacement methods. This crystal structure contains three monomers in the asymmetric unit and was refined to a final *R* factor of 17.5% ($R_{\rm free}$ = 22.1%) at 2.6 Å resolution (Table 1). The final model includes residues Ala10–Leu290 plus a C-terminal α-helix spanning residues Gly304–Glu316 for chain *A*; similarly, Arg11–Ala284 plus the same helix comprising Gly304–Glu316 were traced for monomer *B*. The core region of monomer *C* could be traced from Ala10 to Ala284, but in contrast no electron density was visible for the C-terminal α-helix. This helix displays very high mobility in the other two monomers, as shown by the high average *B* factors (>100 Å^2).

The protein folds into a trimer, with each monomer composed of two distinct domains: an N-terminal domain made up of a central β-sheet of eight strands flanked by three α-helices and a helical C-terminal domain composed of four α-helices (α4–α7; Fig. 1*a*). A β-hairpin formed by strands β4 and β5, inserted between the second (β3) and third (β6) strands of the N-terminal β-sheet, interacts with strand β10 from the neighbouring monomer in the trimer (Fig. 1*b*). The structure of an Rv2714 homologue also of unknown function, CGL1923 from *Corynebacterium glutamicum*, has been solved at 2.35 Å resolution by the Midwest Center for Structural Genomics (PDB entry 2p90). The two structures can be superimposed with a root-mean-square deviation (r.m.s.d.) of 2 Å for the core, excluding the C-terminal region (helix α7 is missing from the CGL1923 structure). The quaternary organization of both proteins in the different crystal forms indicates that the biological unit is most likely a trimer, which is in agreement with dynamic light-scattering measurements made on Rv2714 prior to crystallization (data not shown).

3.2. Two bacterial classes of Rv2714 homologues

The family of homologues detectable from amino-acid sequence comparisons is largely restricted to Actinobacteria. Database sequence searches with *BLASTP* (Altschul *et al.*, 1997) identified 56 nonredundant homologues using a cutoff of $E < 10^{-33}$ after discarding sequences that were too similar (>95% identity in pairwise alignments). A distance tree calculated from the multiple sequence alignment clearly identified two groups or subfamilies of proteins (Fig. 2), the second group being mostly populated by paralogous sequences from the first. This is particularly the case for mycobacteria, as the majority of these species have a homologue of Rv2714 as well as a paralogue (Rv2125 in *M. tuberculosis*), suggesting that the proteins in each similarity cluster might have acquired different biochemical or cellular functions. A notable exception appears to be the *Corynebacterium* genus, as the four sequenced genomes only contained the Rv2714-like gene with no paralogue (Fig. 2).

Both protein groups show different conservation patterns when aligned separately. Within each group, a mycobacterial subgroup is clearly defined, in line with their common descent. The dendrogram suggests that gene duplication was an ancient event in Actinobacteria, given the similar distribution of the paralogues (one per cluster) and the absence of a paralogue pair within the same similarity cluster. This was further reinforced by maximum-likelihood analysis using the *PHYML* program (Guindon & Gascuel, 2003), which produced results that were consistent with those shown in Fig. 2, although marginal differences were found in the clustering of some of the most distant sequences (data not shown).

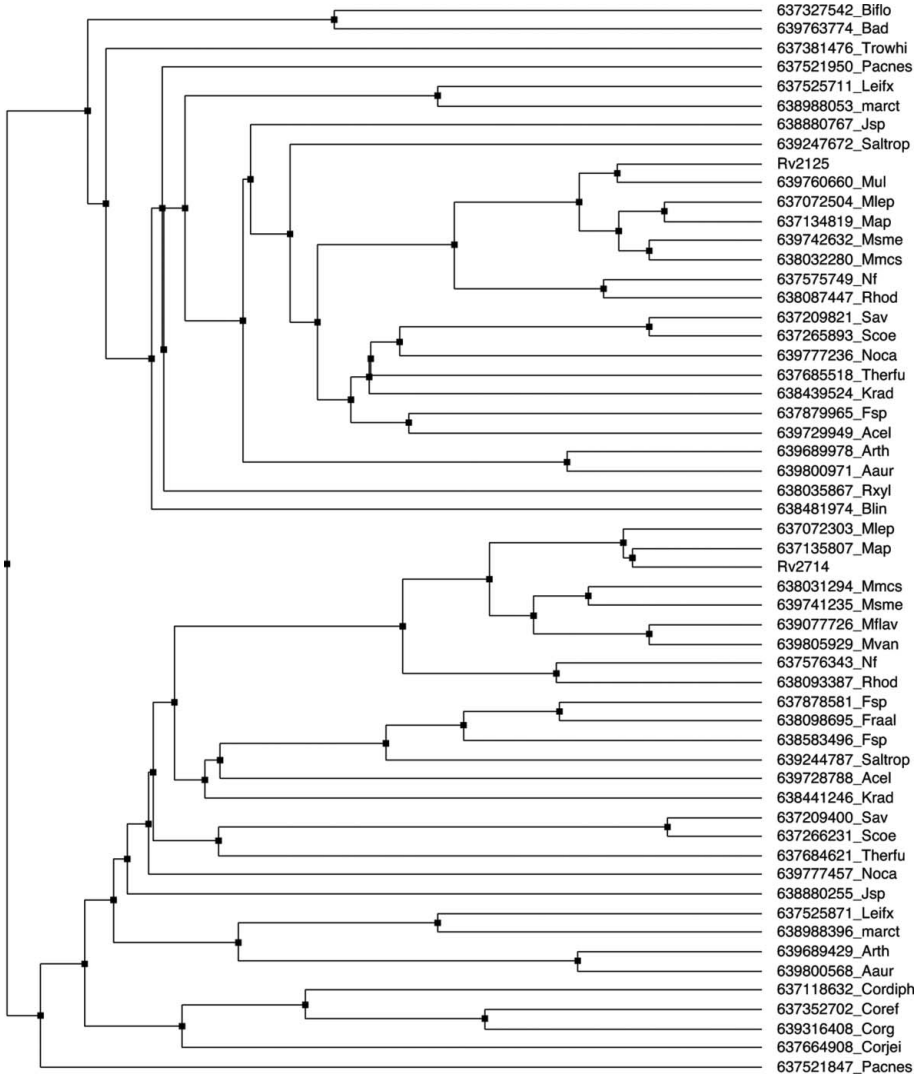

Figure 2
Distance dendrogram (based on pairwise sequence identities) of 56 sequences homologous to Rv2714 reveals two distinct similarity clusters. Sequence identities are in the range 23–86% for the upper cluster and 39–88% in the lower cluster (highly similar sequences with >95% identity were omitted). Each similarity cluster is mostly composed of paralogous sequences from the other. Note that corynebacterial genomes seem to have only one Rv2714-like gene and no paralogue. Except for Rv2125 and Rv2714 from *M. tuberculosis*, all other genes are identified by the Gene Object ID from the Integrated Microbial Genomes database (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi) followed by the abbreviated species name. Abbreviations: Aaur, *Arthrobacter aurescens*; Acel, *Acidothermus cellulolyticus*; Arth, *Arthrobacter* sp.; Bad, *Bifidobacterium adolescentis*; Biflo, *B. longum*; Blin, *Brevibacterium linens*; Cordiph, *Corynebacterium diphtheriae*; Coref, *C. efficiens*; Corg, *C. glutamicum*; Corjei, *C. jeikeium*; Fraal, *Frankia alni*; Fsp, *Frankia* sp.; Jsp, *Janibacter* sp.; Krad, *Kineococcus radiotolerans*; Leifx, *Leifsonia xyli*; Map, *Mycobacterium avium* subsp. *paratuberculosis*; marct, marine actinobacterium PHSC20C; Mflav, *M. flavescens*; Mlep, *M. leprae*; Mmcs, *Mycobacterium* sp. MCS; Msme, *M. smegmatis*; Mul, *M. ulcerans*; Mvan, *M. vanbaalenii*; Nf, *Nocardia farcinica*; Noca, *Nocardioides* sp.; Pacnes, *Propionibacterium acnes*; Rhod, *Rhodococcus* sp; Rxyl, *Rubrobacter xylanophilus*; Saltrop, *Salinispora tropica*; Sav, *Streptomyces avermitilis*; Scoe, *S. coelicolor*; Therfu, *Thermobifida fusca*; Trowhi, *Tropheryma whipplei*.

The multiple sequence alignment within the Rv2714 similarity cluster (Supplementary Fig. 1[1]) reveals about 20 strictly conserved positions and a larger number of positions with conservative substitutions. According to this alignment, the central α/β core (residues 40–230) is more conserved than either the N-terminal segment or the C-terminal helical domain. Moreover, similar amino-acid residues are observed to define the monomer–monomer interfaces (Supplementary Fig. 1[1]), pointing to a conserved trimeric organization for the whole family of homologous proteins.

3.3. The putative ligand-binding pocket

The projection of overall sequence similarity onto the molecular surface of the Rv2714 trimer reveals two clearly distinct sides of the protein; the concave region of the trimer is significantly more conserved in the family than the convex region (Figs. 3*a* and 3*b*). Furthermore, this concave face displays a narrow and deep pocket lined by well conserved residues in each monomer, suggesting that this region could fulfil a functional role as the binding site for a potential ligand (see below). The pocket has a largely hydrophobic character, defined by residues Phe47, Val153, Met155, Ile185, Val187, Pro188, Gly208 and Thr210 from one monomer, and Phe90 and Phe95 from the β-hairpin ($\beta4$–$\beta5$) of a neighbouring monomer (Fig. 3*c*). All of these positions are highly conserved in the Rv2714-like family of homologues, except for Phe95, which can be substituted by other hydrophobic (Trp, Tyr, Leu, Val) residues (Supplementary Fig. 1[1]). In two of the three independent monomers the C-terminal helix $\alpha7$ acts as a 'lid' covering the entrance to the pocket (Fig. 3*c*). The bottom of the pocket contains two polar residues, Asn193 and His168, with the imidazole group of the latter forming a salt bridge to Glu196. The Asn position is highly variable in the multiple sequence alignment, but His168 and Glu196 (two residue types frequently found in the catalytic centre of enzymes) are conserved in the majority of homologous sequences to Rv2714 (Supplementary Fig. 1). Thus, 25 sequences out of 29 contain the His-Glu pair at these positions, while three additional sequences preserve the glutamate but have the histidine replaced by a serine residue.

3.4. Structural comparisons

Several methods were used to identify structural homologues of Rv2714. Excluding *C. glutamicum* CGL1923, these searches returned different results depending on the particular method used. Thus, the program *VAST* (Gibrat *et al.*, 1996) detected the highest number of structural homologues (25 hits containing more than 120 aligned residues aligned with r.m.s.d.s of <3.5 Å). In contrast, *DALI* (Holm & Sander, 1995) and *MATRAS* (Kawabata & Nishikawa, 2000) detected the longest topologically equivalent regions (160–190 residues), partly owing to a better tolerance of insertions and deletions in the structural comparisons, while *SSM* (Krissinel & Henrick, 2004) identified shorter regions with lower r.m.s.d.s than the other methods. Despite these dissimilar results, all methods consistently detected a structural similarity between Rv2714 and several type I purine nucleoside phosphorylases (PNPs) or nucleosidases (Pugmire & Ealick, 2002). Type I PNPs display a conserved protomer arrangement as trimers or hexamers and can accept a wide range of different substrates among purine or pyrimidine nucleotides. However, the monomer–monomer interfaces in the Rv2714 trimer are completely different from those observed in trimeric or hexameric type I PNPs.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: TB5013).

Table 1
Crystallographic data-collection and refinement statistics for Rv2714.

Values in parentheses are for the highest resolution shell.

	SeMet Rv2714	Native Rv2714
Data collection		
Resolution (Å)	53–2.7 (2.85–2.7)	48–2.6 (2.74–2.6)
Wavelength (Å)	0.9793	0.931
Space group	$P6_3$	$P3_221$
Unit-cell parameters (Å)	$a = b = 136.42$, $c = 120.43$	$a = b = 140.49$, $c = 129.77$
Redundancy	7.7 (7.7)	3.7 (3.8)
Completeness (%)	99.9 (100)	99.8 (100)
$R_{\rm merge}$†	0.11 (0.47)	0.07 (0.53)
No. of unique reflections	35001	45795
$\langle I/\sigma(I)\rangle$	16.4 (4.0)	18.1 (2.6)
Refinement statistics		
σ cutoff		0
Resolution (Å)		44–2.6 (2.65–2.6)
No. of reflections used		43446 (2534)
B factors (Å^2)		
Average		50.8
Protein		51.1
Water		40.2
No. of non-H atoms‡		
Protein		6580
Water		233
R.m.s.d. from ideal		
Bond lengths (Å)		0.007
Angles (°)		1.01
$R_{\rm work}$§		0.175 (0.237)
$R_{\rm free}$§		0.221 (0.300)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. ‡ Per asymmetric unit. § $R = \sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}|| / \sum_{hkl} |F_{\rm obs}|$. $R_{\rm free}$ was calculated for 5% of the data not used at any stage of structural refinement.

The equivalent regions between Rv2714-like and PNP enzymes, as defined by *SSM*, comprise between 134 and 156 residues that can be superposed with r.m.s.d.s of 2.6–3.3 Å. The superimposable regions essentially span the protein core (Fig. 4*a*), whereas the segments $\beta4$–$\beta5$ and $\alpha5$–$\alpha7$ from Rv2714 do not have structural equivalents and represent a signature of the Rv2714-like family of homologues. The substrate-binding sites of PNPs, as evaluated from the position of the bound ligand in various PNP crystal structures (Fig. 4*b*), are found within the conserved concave face of the Rv2714 trimer and are located adjacent to, but do not coincide with, the identified pocket in Rv2714 (Fig. 4*b*). In fact, the phosphoribose site present in all PNPs is completely occluded by a significantly larger loop in Rv2714 that connects $\beta2$ to $\alpha1$. The adjacent pocket observed in the mycobacterial protein is instead generated because of a change in the main-chain trace of the loop coming out from strand $\beta7$. The bottom of both deep pockets display substantial differences in terms of the bounding residues: while polar residues are critically located in PNPs to position the nucleophilic phosphate group (Pugmire & Ealick, 2002), in Rv2714 this floor is mostly of hydrophobic character. Nevertheless, since the size and shape of the Rv2714 pocket might be compatible with the fixation of nucleotidic substrates, we soaked several potential PNP substrates/inhibitors in the trigonal crystal form of the protein, in which the putative binding site is exposed to the solvent. Tested ligands included inhibitors that were cocrystallized with one or more PNPs (*e.g.* formycine) as well as potential substrates for PNPs/nucleosidases (*e.g.* various ribonucleosides and deoxyribonucleosides). However, for all soaked compounds (10 m*M* concentration), analysis of Fourier difference maps failed to show supporting electron density for a pocket-bound ligand (data not shown). In addition, our attempts to measure a detectable PNPase activity for Rv2714 using inosine as a substrate (in conditions under which *E. coli* PNPase is fully active) were also negative. Taken together, these results seem to rule out a putative PNP/nucleosidase catalytic activity for the Rv2714 family of proteins.

In addition to PNPs/nucleosidases, three carboxypeptidases were also found to be structurally similar to Rv2714: liver carboxypeptidase from duck (PDB code 1h8l; Aloy *et al.*, 2001), bovine pancreatic carboxypeptidase (PDB code 1m4l; Kilshtain-Vardi *et al.*, 2003) and pancreatic carboxypeptidase from pig (PDB code 1zg7; Adler *et al.*, 2005). Superimposition of the Rv2714 monomer with two of these enzymes crystallized in complex with inhibitors (1zg7 and 1h8l) reveals a common conserved core (Fig. 4*c*). Using *SSM*, 132 residues from 1h8l were found to occupy equivalent positions with an r.m.s.d. of 2.7 Å, while for 1zg7 126 equivalent positions were superposed with an r.m.s.d. of 2.93 Å. Interestingly, the active sites of both carboxypeptidases (considered to be the inhibitor-binding sites) perfectly coincide with the conserved pocket in Rv2714 (Fig. 4*d*). However, a carboxypeptidase activity for Rv2714 is unlikely for several reasons. Firstly, carboxypeptidases are metalloenzymes containing a catalytic Zn atom and no metal-binding site was observed in the structure of Rv2714. Secondly, the active site of carboxypeptidases is a shallow solvent-accessible crevice (consistent with peptide substrates), while the pocket in Rv2714 is deeper and is occluded by the α7 helix (Fig. 1), suggesting that it could accommodate relatively small molecules. Thirdly, the electrostatic properties of the Rv2714 pocket (largely hydrophobic) are clearly distinct from those observed for the active site of carboxypeptidases.

Clear structural homology was also detected with the chloroplast group II intron splicing factor CRS2 (PDB code 1ryn; Ostheimer *et al.*, 2005) and hence with a functionally divergent group of bacterial enzymes displaying peptidyl-tRNA hydrolytic activity and considered to be evolutionary ancestors of CRS2 (Ostheimer *et al.*, 2005). In particular, structural similarity was detected between Rv2714 and the peptidyl-tRNA hydrolase (PTH) from *M. tuberculosis* (PDB code 2z2i; Selvaraj *et al.*, 2007). However, in addition to a different quaternary structure organization, the ligand/substrate sites in the related CRS2 and PTHs differ radically from the putative active pocket in Rv2714. Taken together, the above observations strongly suggest a new functional, probably enzymatic, branch for the Rv2714-like family of actinobacterial proteins.

4. Concluding remarks

The structure of Rv2714 reveals a protein core similar to that observed in PNPs/nucleosidases and carboxypeptidases, with a highly conserved pocket that could represent a binding site for an as yet unidentified substrate or ligand. However, our results appear to rule out a possible carboxypeptidase or PNP/nucleosidase function for Rv2714. From a structural point of view, the putative Rv2714

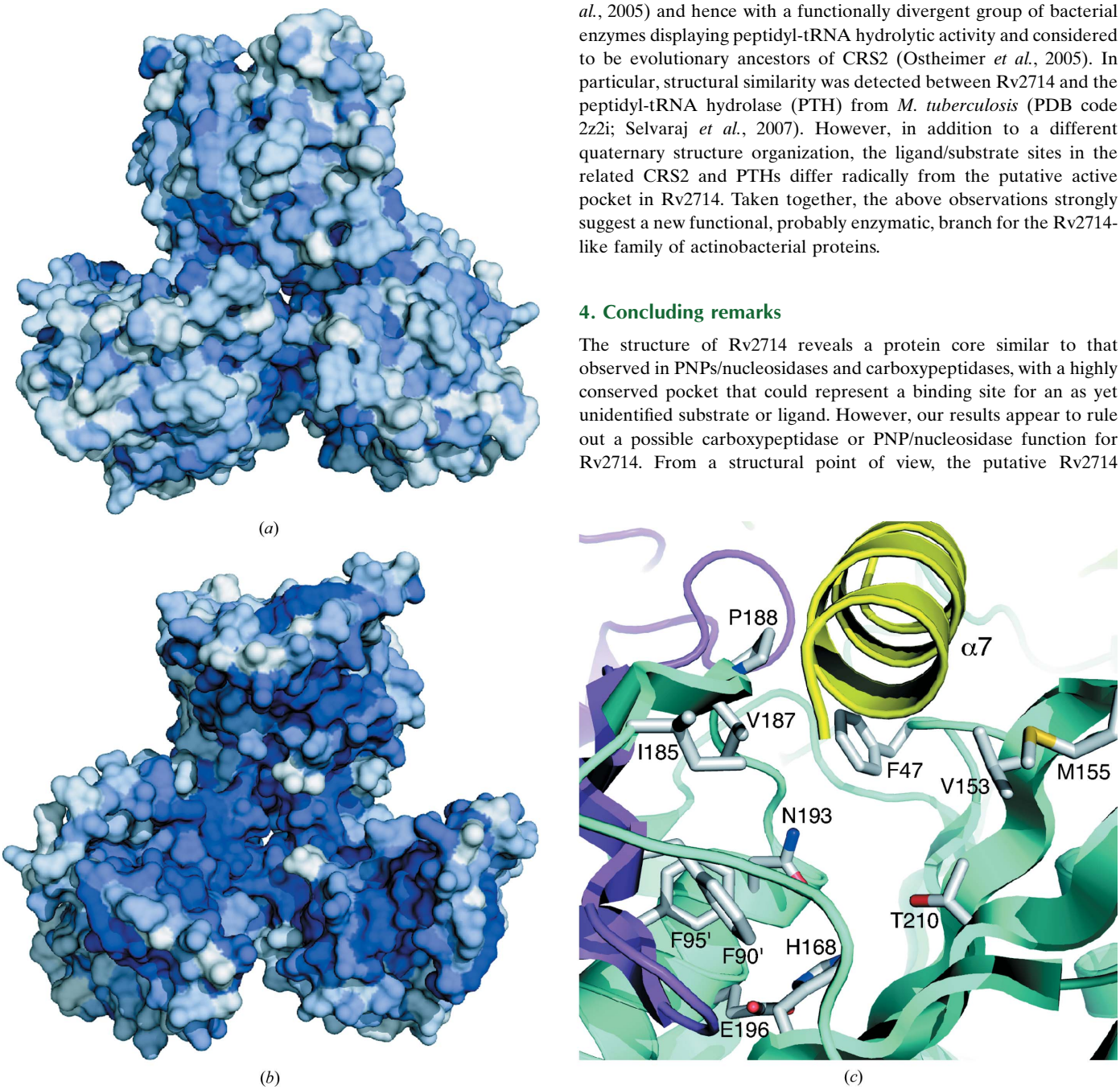

Figure 3
The convex (*a*) and concave (*b*) faces of the Rv2714 trimer are coloured according to amino-acid conservation (dark blue, conserved; white, variable). (*c*) Putative ligand-binding site of Rv2714. Conserved residues defining the pocket are indicated (see text). The entrance to the site is partially occluded by helix α7 (in yellow).

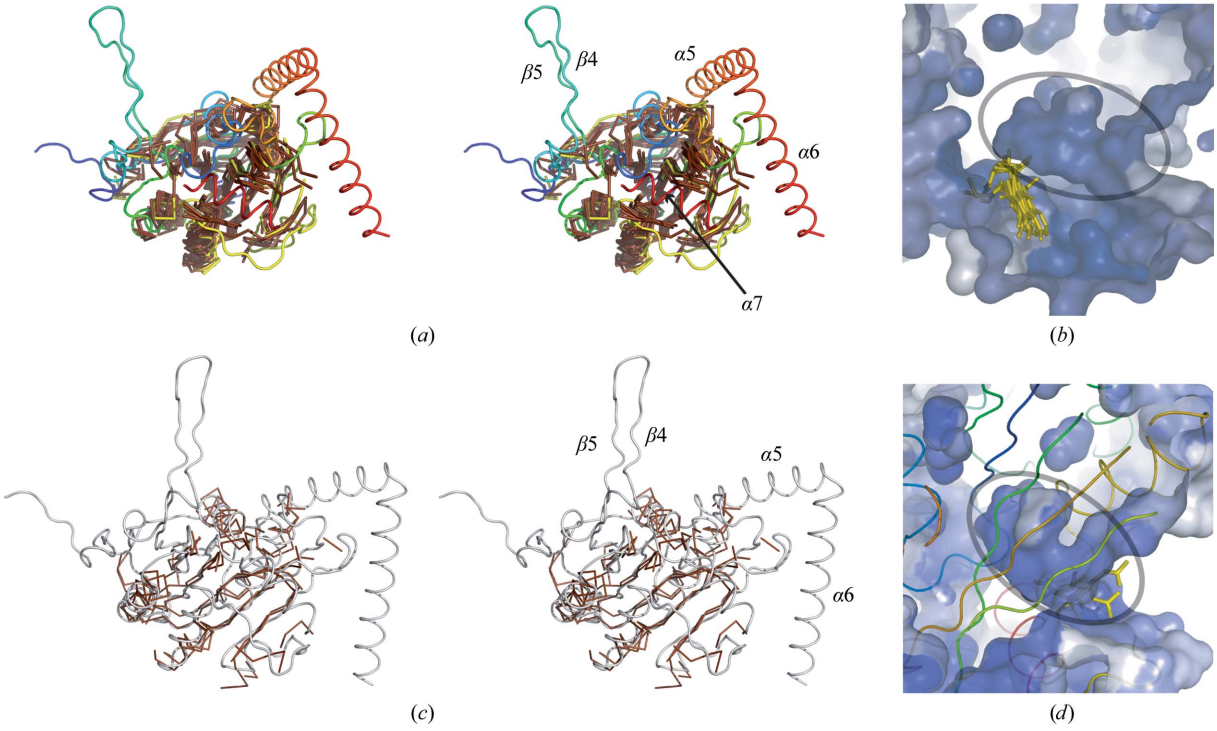

Figure 4
Structural homologues of Rv2714. (*a*) Stereoview showing the structural superposition of Rv2714, colour-coded from blue (N-terminus) to red (C-terminus), with nine PNPs/nucleosidases that were crystallized in complex with different substrates or inhibitors (PDB codes 1a9t, 1b8o, 1nc3, 1rxc, 1z5n, 1zos, 2a8y, 2bsx and 2h8g). Structurally equivalent regions of the PNPs/nucleosidases are shown in brown. (*b*) The molecular surface of Rv2714, colour-coded according to residue conservation (see Fig. 3), showing the position of the bound ligands in PNPs/nucleosidases (yellow sticks). The putative binding pocket in Rv2714 is circled. (*c*) Stereoview showing the structural superposition of Rv2714 (light grey) with two carboxypeptidases (PDB codes 1h8l and 1zg7; structurally equivalent regions are shown in brown). For clarity, the view is slightly rotated from that shown in (*a*). (*d*) Molecular surface (colour-coded as in *b*) and C^{α} trace of Rv2714. The position of the bound ligands in the two carboxypeptidase structures (yellow sticks) coincide with the putative binding pocket of Rv2714 (circled).

substrate-binding pocket is more hydrophobic than a typical PNP, suggesting that the actual substrate (ligand) could be different from standard nucleosides. Moreover, the $\beta4$–$\beta5$ and $\alpha5$–$\alpha7$ subdomains are only found in Rv2714-like bacterial proteins and constitute specific signatures of this protein family. These structural elements may play a functional role, since the $\beta4$–$\beta5$ hairpin serves as an anchor domain for trimer assembly (differing from that observed in PNPs), while the C-terminal helices $\alpha5$–$\alpha7$ provide a particular environment to the conserved pocket in which $\alpha7$ acts as a lid to the putative binding site. Further experimental work is required to validate these functional hypotheses and narrow down the class of substrates (or ligands) that could be specific for the Rv2714 family of actinobacterial proteins.

We are grateful to W. Parker and S. Ealick for performing enzymatic assays and for insightful discussions and to V. Bondet and J. Bellalou for help in protein production. We acknowledge the ESRF for provision of synchrotron-radiation facilities and the ESRF staff for assistance in X-ray data collection. This work was partially supported by grants from Institut Pasteur, the CNRS (France) and the European Union FP6 Project NM4TB (contract No. LSHP-CT-2005-018923). MG acknowledges partial support from the Fondo Clemente Estable (Uruguay), contract 2007-377.

References

Adler, M., Bryant, J., Buckman, B., Islam, I., Larsen, B., Finster, S., Kent, L., May, K., Mohan, R., Yuan, S. & Whitlow, M. (2005). *Biochemistry*, **44**, 9339–9347.
Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *Acta Cryst.* D**61**, 850–855.
Aloy, P., Companys, V., Vendrell, J., Aviles, F. X., Fricker, L. D., Coll, M. & Gomis-Ruth, F. X. (2001). *J. Biol. Chem.* **276**, 16177–16184.
Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Cole, S. T. (2002). *Eur. Respir. J. Suppl.* **36**, 78s–86s.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gibrat, J. F., Madej, T. & Bryant, S. H. (1996). *Curr. Opin. Struct. Biol.* **6**, 377–385.
Guindon, S. & Gascuel, O. (2003). *Syst. Biol.* **52**, 696–704.
Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Kawabata, T. & Nishikawa, K. (2000). *Proteins*, **41**, 108–122.
Kilshtain-Vardi, A., Glick, M., Greenblatt, H. M., Goldblum, A. & Shoham, G. (2003). *Acta Cryst.* D**59**, 323–333.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Ostheimer, G. J., Hadjivassiliou, H., Kloer, D. P., Barkan, A. & Matthews, B. W. (2005). *J. Mol. Biol.* **345**, 51–68.
Pugmire, M. J. & Ealick, S. E. (2002). *Biochem. J.* **361**, 1–25.
Selvaraj, M., Roy, S., Singh, N. S., Sangeetha, R., Varshney, U. & Vijayan, M. (2007). *J. Mol. Biol.* **372**, 186–193.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215–230.
Wingfield, P. (2000). *Current Protocols in Protein Science*, edited by J. Coligan, B. Dunn, H. Ploegh, D. Speicher & P. Wingfield, Vol. I, pp. 5.3.9–5.3.14. New York: Wiley.